HHS Public Access

Author manuscript

J Chem Theory Comput. Author manuscript; available in PMC 2024 August 13.

Published in final edited form as:

J Chem Theory Comput. 2024 February 13; 20(3): 1051–1061. doi:10.1021/acs.jctc.3c00993.

Enhanced Sampling of Buried Charges in Free Energy Calculations using Replica Exchange with Charge Tempering

Xiaorong Liu[†], **Charles L Brooks III**[†,‡]

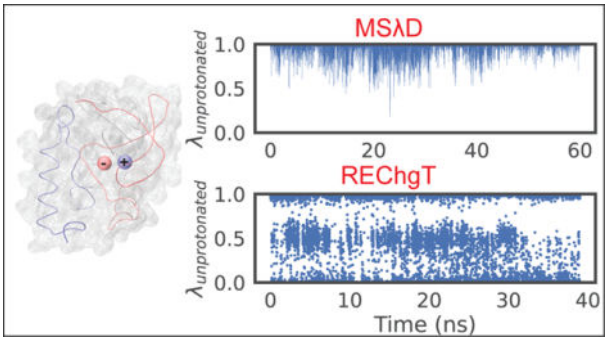
[†]Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109, USA

[‡]Biophysics Program, University of Michigan, Ann Arbor, Michigan 48109, USA

Abstract

Buried ionizable groups in proteins often play important structural and functional roles. However, it is generally challenging to study the detailed molecular mechanisms solely based on experimental measurements. Free energy calculations using atomistic simulations, on the other hand, complement experimental studies and can provide high temporal and spatial resolution information that leads to mechanistic insights. Nevertheless, it is also well recognized that sufficient sampling of such atomistic simulations can be challenging, considering that structural changes related to the buried charges may be very slow. In the present study, we describe a simple but effective enhanced sampling technique called replica exchange with charge tempering (REChgT) with a novel free energy method, multisite λ dynamics (MSλD), to study two systems containing buried charges, pKa prediction of a small molecule, orotate, in complex with the dihydroorotate dehydrogenase (DHOD), and relative stability of a Glu-Lys pair buried in the hydrophobic core of two variants of Staphylococcal nuclease (SNase). Compared to the original MSλD simulations, the usage of REChgT dramatically increases sampling in both conformational and alchemical space, which directly translates into a significant reduction of wall time to converge the free energy calculations. This study highlights the importance of sufficient sampling towards developing improved free energy methods.

Graphical Abstract



brookscl@umich.edu .

Supporting Information Available

Introduction

Although it is inherently unfavorable to bury ionizable groups in the hydrophobic core of proteins due to the large desolvation free energy penalty, buried ionizable groups in proteins play important functional roles in many biological processes, such as catalysis,[1–4] translocation of metal ions[5] and protons,[6–9] and electron transport.[10] Understanding the properties of these buried charges and dominant factors of stabilization is important in establishing structure-function relationships in these proteins, and could further guide the rational design of novel enzymes. However, detailed molecular mechanisms solely based on experimental measurements are often difficult to discover. High spatial and temporal resolution information, such as protonation state of the ionizable groups, protein conformational dynamics, and fluctuations of penetrated water molecules (if present), are often required to elucidate the physical basis of how buried ionizable groups are stabilized to exert critical functions.

Free energy simulations using explicit solvent all-atom force fields, on the other hand, can complement experimental studies and provide the required structural and dynamic information that leads to mechanistic insights. For example, thermodynamic integration (TI) has been applied to study a series of Staphylococcal nuclease (SNase) variants with buried Glu-Lys pairs.[11,12] Constant pH molecular dynamics using multisite λ dynamics (MSλD) has been utilized to comprehensively investigate the buried ionizable residues of the matrix-2 (M2) protein from influenza A virus.[13] Important insights have been obtained from these free energy simulations to better understand the molecular-level functioning of these proteins. However, obtaining sufficient sampling can sometimes be challenging for systems with buried charges, as reported previously in numerous studies.[11,14–16] In many cases, substantial rearrangement of the local structure is involved, and sufficient sampling of such conformational changes is required to reliably estimate the corresponding free energy values. Therefore, we need to either perform extensive standard molecular dynamics simulations or combine enhanced sampling techniques into free energy simulations of such challenging systems.

Many enhanced sampling methods have been developed and applied to free energy simulations, such as different variants of replica exchange methods including replica exchange with solute tempering,[17–21] grand canonical Monte Carlo,[22–25] Gaussian accelerated molecular dynamics,[26–29] weighted ensemble,[30–32] metadynamics,[33–36] orthogonal space tempering[37–40] and expanded ensemble.[41–43] For more complete reviews, we refer the reader to references.[44–49] Despite significant advancements in enhanced sampling methods, sufficient sampling of systems with buried charges in many cases remains challenging. In the present work, we have carefully analyzed MSλD free energy simulations of a small molecule, orotate, in complex with the dihydroorotate dehydrogenase (DHOD) (Figure 1A), and found that local conformations related to the buried ionizable carboxylate group of orotate are kinetically trapped and undergo very slow transitions, which creates a sampling bottleneck for free energy calculations aiming to predict the pKa of orotate. To increase sampling efficiency of MSλD simulations in such systems, we propose a simple but effective approach called replica exchange with charge tempering (REChgT), where partial charges of residues related to the local interaction network are

reduced to facilitate necessary structural rearrangement, and replica exchange is used to recover the correct ensemble. Applying this strategy to study an even more challenging system, the relative stability of a Glu-Lys pair buried in the hydrophobic core of two variants of SNase (Figure 1B), suggests that the usage of REChgT can significantly increase sampling as well as increase the convergence rate of free energy calculations. This work highlights the importance of sufficient sampling towards developing improved free energy methods.

Methods

MSλD methodology

The MSλD methodology has been reported previously, and we refer the readers to references[50–53] for more details. In brief, the potential energy of the system is shown in Eq. 1,

$$V = V(x_0, x_0) + \sum_{s=1}^{M}\sum_{i=1}^{N_s} \lambda_{si}(V(x_0, x_{si}) + V(x_{si}, x_{si})) + \sum_{s=1}^{M}\sum_{i=1}^{N_s}\sum_{t=s+1}^{M}\sum_{j=1}^{N_t} \lambda_{si}\lambda_{tj}V(x_{si}, x_{tj}) + V_{bias}\left(\left\{\lambda\right\}\right) \quad (1)$$

where x_0 and x_{si} are coordinates of the environment and substituent i at site s, respectively. M is the number of sites and N_s is the number of substituents at site s. $V_{bias}(\{\lambda\})$ is the biasing potential that is critical for efficient sampling in alchemical space and is a function of λ's only.[52,53] As shown in Eqs. 2–6 below, there are four types of biasing potentials utilized in the latest framework of MSλD,

$$V_{bias}(\{\lambda\}) = V_{fixed}(\{\lambda\}) + V_{quadratic}(\{\lambda\}) + V_{end}(\{\lambda\}) + V_{skew}(\{\lambda\}) \quad (2)$$

$$V_{fixed}\left(\left\{\lambda\right\}\right) = \sum_{s}^{M}\sum_{i}^{N_s} \phi_{si}\lambda_{si} \quad (3)$$

$$V_{quadratic}\left(\left\{\lambda\right\}\right) = \sum_{s}^{M}\sum_{i}^{N_s}\sum_{j>i}^{N_s} \psi_{si,sj}\lambda_{si}\lambda_{sj} \quad (4)$$

$$V_{end}\left(\left\{\lambda\right\}\right) = \sum_{s}^{M}\sum_{i}^{N_s}\sum_{j\neq i}^{N_s} \omega_{si,sj}\lambda_{si}\lambda_{sj}/(\lambda_{si} + 0.017) \quad (5)$$

$$V_{skew}\left(\left\{\lambda\right\}\right)=\sum_{s}^{M}\sum_{i}^{N_s}\sum_{j\neq i}^{N_s}\chi_{si,sj}\lambda_{sj}\left(1-e^{-\lambda_{si}/0.18}\right) \quad (6)$$

where $\{\phi_{si}\}$, $\{\psi_{si}\}$, $\{\omega_{si,sj}\}$ and $\{\chi_{si,sj}\}$ are biasing potential parameters. The linear function $V_{fixed}(\{\lambda\})$ is to make the free energies at end states similar, $V_{quadratic}(\{\lambda\})$ is designed to remove the quadratic component of free energy barriers for transitions between two end states, $V_{end}(\{\lambda\})$ is used to avoid trapping at the deep free energy minimum near the end state, and $V_{skew}(\{\lambda\})$ is an additional term to improve the fitting of free energy profiles. Note that in charging free energy calculations, the medium usually responds linearly to the introduction of charges, so the free energy varies predominantly quadratically with λ. Therefore, the free energy profiles in this work can be well approximated by the summation of $V_{fixed}(\{\lambda\})$ and $V_{quadratic}(\{\lambda\})$ (see Results and discussion for more details).

Free energy estimation

For MSλD simulations, we collect the λ trajectories and calculate free energies using the estimator in Eq. 7.[51]

$$\Delta G^{MS\lambda D}(\{\lambda_{si}\}\rightarrow\{\lambda_{sj}\})\approx -k_BTln\frac{P(\{\lambda_{sj}\}>0.99)}{P(\{\lambda_{si}\}>0.99)}-[V_{bias}(\{\lambda_{sj}\}=1)-V_{bias}(\{\lambda_{si}\}=1)] \quad (7)$$

For REChgT simulations where replica exchange was performed among thermodynamic states/conditions with different charge scaling factors (see Free energy simulations of DHOD/orotate below for details), we collect all conformations under the condition of charge scaling factor of 1.0 and use the estimator in Eq. 7 to compute relative free energies. However, for 2D REChgT simulations where exchange was attempted between replicas with different charge scaling factors and $V_{bias}(\{\lambda\})$ (see section Free energy simulations of SNase mutants for details), we utilized the binless weighted histogram analysis method (WHAM)[54,55] to unbias each snapshot and compute relative free energies. Specifically,

$$f_i=-ln\sum_{j=1}^{K}\sum_{n=1}^{N_j}\frac{exp\left(-\beta V^{i}_{total\ bias}(X_{jn})\right)}{\sum_{k=1}^{K}N_k exp\left(f_k-\beta V^{k}_{total\ bias}(X_{jn})\right)} \quad (8)$$

$$<P\left(\eta\right)>=\sum_{j=1}^{K}\sum_{n=1}^{N_j}\frac{\delta(\eta(X_{jn})-\eta)}{\sum_{k=1}^{K}N_k exp\left[f_k-\beta V^{k}_{total\ bias}(X_{jn})\right]} \quad (9)$$

where K is the number of replicas. N_j is the number of snapshots in condition j. $V^k_{total\ bias}(X_{jn})$ is the total biasing potential energy computed at condition k for snapshot n sampled at condition j, which has two components: $V_{bias}(\{\lambda\})$ as shown in Eq. 2 and the change in electrostatic interactions due to charge scaling. Although it is mathematically equivalent to use the total potential energies in Eqs. 8 and 9, the use of $V_{total\ bias}$ greatly reduces the computational cost associated with recomputing energies for sampled snapshots. η was defined such that a given snapshot is categorized as either one of the end states (i.e., with λ cutoff 0.99) or an intermediate state.

Preparation of the DHOD/orotate system

The crystal structure of E. coli DHOD protein in complex with small molecule orotate was obtained from the RCSB protein data bank (PDB ID: 1F76)[56] and used as our initial structure. The cofactor FMN and water molecules in the crystal structure were also kept. The N- and C-termini of DHOD were capped with an acetyl group and N-methyl amide, respectively. The system was solvated in a cubic water box, with at least 12 Å from the protein to the nearest box edge. A proper amount of Na^+ and Cl^- ions were added to bulk water using the MMTSB tool set[57] to neutralize the system and to mimic the physiological condition of 150 nM NaCl.

For the MSλD simulations, we need to construct a hybrid orotate molecule where all atoms belonging to the maximum common substructure (MCS) are represented once and atoms unique to either the protonated or unprotonated state are represented explicitly. The identification of the MCS and assignment of partial charge to each atom based on the charge renormalization scheme were achieved through the automated workflow *msld_py_prep*.[58] In conventional MSλD simulations and the free energy perturbation (FEP) like windowing approach, a two-state model (i.e., one protonated state and one unprotonated state) was used for the hybrid orotate molecule. In REChgT simulations, the hybrid orotate was represented by a three-state model that considers proton tautomerism, i.e., including one unprotonated state and two protonated states. Inclusion of such proton tautomerism is expected to facilitate sampling of the protonated states, but have little impact on the timescale of forming or breaking hydrogen bonds (HBs). Therefore, it should not affect the comparison of sampling efficiency of the rate-limiting structural reorganization, that is, rearrangement of the local HB interaction network (see Figure 4).

Preparation of the SNase systems

Initial structures of the V23K/L36E and V23E/L36K mutants of the highly stable SNase variant, Δ+PHS, were obtained from PDB with IDs of 6AMF[59] and 3NHH,[60] respectively. Coordinates of water molecules and ions from the crystal structures were preserved. The N- and C-termini of SNase were capped with an acetyl group and N-methyl amide, respectively. PROPKA3[61,62] was used to determine the protonation states of titratable residues, and no residue was required to change its protonation state at pH 7, the pH representing the conditions of our simulated solution. The Glu-Lys pair at positions 23 and 36 of SNase were treated as hybrid residues in all of our simulations. The Glu and Lys residues were modeled with three and two states, respectively. The protein was solvated in a cubic water

box containing ~ 12,700 and 13,800 TIP3 water molecules for V23E/L36K (3NHH) and V23K/L36E (6AMF), respectively. This allowed for at least 12 Å between the protein and the nearest box edge. A proper amount of Na^+ and Cl^- ions were randomly placed in bulk water using the MMTSB tool set[57] to neutralize the system and to mimic the physiological condition of 150 nM NaCl.

Details of molecular dynamics simulations

All simulations were performed using pyCHARMM,[63] a newly developed Python framework that embeds CHARMM functionalities and readily integrates with other Python functions and toolkits. The CHARMM/BLaDE platform[64] was used to accelerate simulations on GPUs. Periodic boundary conditions were used in all simulations. All systems were modeled with CHARMM36m force field.[65] The forces of van der Waals interactions were smoothly switched to zero at 12 Å starting from 10 Å. The particle mesh Ewald method[66] was used to treat the long-range electrostatic interactions. SHAKE[67] was applied to constrain any bond involving hydrogen atoms, which allowed for a time step of 2 fs to integrate the equations of motion. Each system was first energy minimized to remove potential steric clashes. Short equilibration simulations at 298.15 K under NVT conditions were then performed for ~ 1 ns, with the force constant of harmonic restrains on protein Cα positions slowly reduced from 5 kcal/mol/Å^2 to zero. NPT simulations at 298.15 K and 1 atm were carried out to allow the whole system to relax without any restraint in the system. The average box volume of the system during this NPT equilibration simulation was used to determine the box size of final production NVT simulations at 298.15 K.

Free energy simulations of DHOD/orotate

To obtain the best possible $V_{bias}(\{\lambda\})$ for MSλD simulations of DHOD/orotate, an FEP-like windowing approach was used to compute the free energy profile as a function of λ. Specifically, 11 windows were used with fixed λ values that were equally spaced between 0 and 1. Each window was simulated for 18 ns, and three independent runs were performed. FastMBAR[68,69] was used to compute free energy as a function of λ. The calculated free energy profiles were then fitted with function $G = a\lambda + b\lambda(1-\lambda)$, where a and b are biasing potential parameters for the fixed and quadratic terms, respectively. Average values of a and b over these three runs were used as the optimal parameters of $V_{bias}(\{\lambda\})$ in all later MSλD and REChgT simulations. Since the summation of $V_{fixed}(\{\lambda\})$ and $V_{quadratic}(\{\lambda\})$ approximated the free energy profile very well (see Figure S1), no endpoint or skew terms were used. We note that these simulations also provide an independent check on the subsequent REChgT calculations.

For the MSλD simulations, five independent runs were performed with different initial velocities. Each simulation lasted for 60 ns. For the REChgT simulations, partial charges of residues Lys 66, Asn 111, Ser 175 and Asn 177 of DHOD were scaled. These four residues interact strongly with the carboxylate group of orotate. Six replicas were used with charge scaling factors of 1.0000, 0.9018, 0.8132, 0.7333, 0.6613 and 0.5963, respectively. This allowed for sufficient weakening of strong local interactions while providing a reasonable

exchange acceptance ratio (~ 0.35 on average). Here, three independent runs were carried out, each for 40 ns/replica.

Free energy simulations of SNase mutants

We first used adaptive landscape flattening (ALF)[52,53] to optimize $V_{bias}(\{\lambda\})$ for MSλD simulations of two SNase mutants. 50 iterations of 100-ps simulations were performed during the first phase of optimization, followed by 11 or 45 iterations of 1-ns simulations during the second phase for V23E/L36K (3NHH) and V23K/L36E (6AMF), respectively. Although it was unclear how many iterations we should run to make $V_{bias}(\{\lambda\})$ sufficiently optimal, the biasing parameters appeared to be converged at the end of our ALF procedure. For each system, the same $V_{bias}(\{\lambda\})$ obtained from the above ALF protocol was used in all of our production simulations using either the original MSλD or the REChgT method.

For the original MSλD simulations, we performed five independent runs, each lasting for 60 ns. For REChgT simulations, we scaled the partial charges of all atoms in residues 23 and 36, i.e., the Glu-Lys pair. Six replicas were used with charge scaling factors of 1.0000, 0.9807, 0.9614, 0.9421, 0.9228, and 0.9035, respectively. Again, this was designed to allow the salt bridge between the Glu-Lys pair to break while facilitating sufficient exchange between neighboring replicas. The average exchange acceptance ratio was ~ 0.69 for V23K/L36E (6AMF) and 0.56 for V23E/L36K (3NHH). Each REChgT simulation was 30 ns/replica. For the 2D REChgT simulations, we had four replicas in the dimension of charge scaling, with scaling factors of 1.0000, 0.9678, 0.9357, and 0.9035, respectively. In the dimension of $V_{bias}(\{\lambda\})$, only $V_{fixed}(\{\lambda\})$ was varied (see Results and discussion for more details), with the fixed biasing parameters equally spaced with a spacing of 1.5 kcal/mol. For 3NHH, the fixed biasing parameters were −4.5 kcal/mol to 3.0 kcal/mol away from the original values obtained from ALF, which gave six replicas in the $V_{bias}(\{\lambda\})$ dimension and 24 replicas in total. For 6AMF, the fixed biasing parameters were −6.0 kcal/mol to 4.5 kcal/mol away from the original values obtained from ALF, which gave eight replicas in the $V_{bias}(\{\lambda\})$ dimension and 32 replicas in total. Each 2D REChgT simulation was 100 ns/replica, and the average exchange acceptance ratios were 0.55 and 0.54 for V23K/L36E (6AMF) and V23E/L36K (3NHH), respectively.

To compute the relative stability of SNase mutants when the Glu-Lys pair is changed from the charge-neutral form to the ionized form, we also performed MSλD simulations of Glu and Lys dipeptides in water. This corresponds to the unfolded states in the thermodynamic cycle (see Figure 2). Here we assume that the unfolded state does not form any stable structure such that the effect of protonation/deprotonation is primarily local and can be captured by simulating the corresponding residue in bulk water. Again, the hybrid Glu dipeptide was modeled with three states by considering proton tautomerism, and the hybrid Lys dipeptide was treated with a two-state model. For each dipeptide, we ran five independent MSλD simulations and each lasted for 10 ns.

To validate our λ dynamics simulation results, we also adopted a windowing approach to simulate the two SNase mutants. Here, both residues of the Lys-Glu pair were treated with a two-state model, and $\lambda^{K}_{protonated}$ and $\lambda^{E}_{unprotonated}$ were kept the same in each window. There were

19 windows for each simulation, with λ values of 0.0, 0.1, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95 and 1.0, respectively. Each window lasted for 180 ns, and two independent runs were carried out for each SNase mutant. Free energies were estimated using FastMBAR.[68,69]

If we define $\Delta\Delta G_{fold}(N \rightarrow C)$ as the change in folding free energy when the Glu-Lys pair changes from the neutral form to ionized form, then, based on Figure 2, we can see that:

$$\begin{aligned}
&\Delta\Delta G_{fold}\left(N \rightarrow C\right) \\
&= \Delta G_{fold}\left(U^C\left(KH^+ \cdot E^-\right) \rightarrow F^C\left(KH^+ \cdot E^-\right)\right) - \Delta G_{fold}\left(U^C\left(KH^+ \cdot E^-\right) \rightarrow F^N\left(K \cdot EH\right)\right) \\
&= \Delta G^C_{fold}\left(KH^+ \cdot E^-\right) - \left(\Delta G^N_{fold}\left(K \cdot EH\right) - \Delta G^U\left(N \rightarrow C\right)\right) \\
&= \left(\Delta G^F\left(N \rightarrow C\right) - \Delta G^U\left(N \rightarrow C\right)\right) + \Delta G^U\left(N \rightarrow C\right) \\
&= \left(\Delta G^F\left(N \rightarrow C\right) - \Delta G^U\left(N \rightarrow C\right)\right) + \left(ln\,10 k_B T pK_a^{Glu} - ln\,10 k_B T pK_a^{Lys}\right) \\
&= \left(\Delta G^F\left(N \rightarrow C\right) - \Delta G^U\left(N \rightarrow C\right)\right) + 2.303 k_B T\left(4.4 - 10.4\right)
\end{aligned} \tag{10}$$

Note that $\Delta G^F\left(N \rightarrow C\right) - \Delta G^U\left(N \rightarrow C\right)$ in Eq. 10 can be calculated from our simulations. $\Delta G^U\left(N \rightarrow C\right)$ is to account for the protonation states of Lys and Glu under neutral pH in the unfolded state (i.e., bulk water here), which can be roughly estimated using the definition of the pKa for Lys and Glu residues.[11]

Results and discussion

Slow conformational transition could be rate-limiting in MSλD free energy simulations

In MSλD simulations, the alchemical parameters, λ's, are treated as dynamic variables and spontaneously fluctuate between 0 and 1. Such reversible transitions in λ space are essential for free energy estimations, for example, when using the estimator in Eq. 7. As discussed in previous work,[52,53] the biasing potential $V_{bias}(\{\lambda\})$ (see Eqs. 1–6) is used to flatten the alchemical free energy landscape, thus critical for obtaining efficient sampling in MSλD simulations. However, we also noticed that in challenging cases, even an optimal $V_{bias}(\{\lambda\})$ is not sufficient to generate fast reversible transitions in alchemical space, a necessary condition to obtain well-converged free energy results rapidly. For example, here we carried out MSλD simulations of a small molecule, orotate, in complex with the protein DHOD to compute the relative free energies when the carboxylate group of orotate is either unprotonated or protonated. This is an essential step in free energy based methods to predict the pKa of orotate in the DHOD receptor. We first tried to run many iterations of ALF, an automatic and well-established protocol, to flatten the alchemical free energy landscape and obtain an optimal $V_{bias}(\{\lambda\})$. Unfortunately, reversible transitions in λ-space were rarely observed in all trial runs (data not shown). To obtain the best possible $V_{bias}(\{\lambda\})$, we used an FEP-like windowing approach together with the multistate Bennett acceptance ratio estimator (MBAR)[55,68] to compute the free energy profile as a function of λ, and fitted the free energy profile to obtain optimal parameters for $V_{bias}(\{\lambda\})$ (see Figure S1). As illustrated in Figure 3A, although the orotate molecule should sample both protonated and unprotonated states equally if the simulation is infinitely long based on thermodynamic

principles, it is kinetically trapped in either state throughout 60 ns of MSλD simulations. This demonstrates that an optimal $V_{bias}(\{\lambda\})$ alone is not sufficient for efficient sampling in alchemical space. Further examination of the MSλD simulation trajectories suggests that the carboxylate group of orotate is forming a complex HB interaction network with neighboring residues. Change in protonation state of orotate is tightly coupled to changes in such an interaction network. For instance, when orotate is unprotonated (run 5 in Figure 3A), the carboxylate group acts as an HB acceptor and interacts with the Nδ atoms of Asn 177 and/or Asn 111 (run 5 in Figure 3B), while protonated orotate (run 1 in Figure 3A) may act as a hydrogen bond donor and interact with the Oδ atom of Asn 111 (run 1 in Figure 3B). Breaking and reforming different HBs occurs slowly (at least tens of ns, as illustrated in Figure 3B), which causes slow transitions in alchemical space as well as slow convergence of free energy calculations.

Improved sampling and more rapid free energy calculation in REChgT simulations of DHOD/orotate

To accelerate this type of local structural rearrangement, we designed a simple but effective sampling method called replica exchange with charge tempering (REChgT). In this approach, partial charges of residues related to the local interaction network are reduced to weaken the interaction strength, thus facilitating necessary structural rearrangement. Here, we hypothesize that for systems with buried charges, many local strong interactions related to the buried charge are likely polar, such as HBs and salt bridges. Therefore, reducing electrostatic interactions by reducing the partial charges is probably one of the most effective and simplest approaches to weaken such interactions, thus lowering the barrier of structural reorganization. Furthermore, by incorporating replica exchange into this approach, it promotes mixing of conformations and at the same time allows us to recover the correct ensemble of no charge reduction. Applying REChgT to study the above orotate/DHOD system shows that the local structure is no longer trapped and more structural fluctuations can be observed compared to the original MSλD simulations (Figures 4A and B). For instance, for the HB formed between the oxygen atoms of the orotate carboxylate group and the Oδ atom of Asn 111, the original MSλD simulations could only sample either HB-formed or HB-broken states in a single simulation, while both states can be sampled in one REChgT simulation (Figures 4B). Although these distance probability distributions for atom pairs involved in the HB interaction network are not fully converged even in the REChgT simulations, as shown in Figures 4B, such improved sampling in conformational space appears to be sufficient to be translated into more exhaustive sampling in alchemical space (Figure 4C) and faster convergence of free energy calculations. As shown in Figure S4, two of our REChgT simulations require less than 10 ns to reach a self-consistent free energy estimation, which can not be achieved in 60 ns using the original MSλD simulations. Further comparison of free energy values obtained from REChgT simulations and the windowing approach with the MBAR estimator shows that the two approaches give the same results within uncertainty (Table 1), suggesting that REChgT does not introduce any bias in free energy calculations.

Improved sampling and faster convergence of free energy calculation in 2D REChgT simulations of a more challenging system: SNase

To examine how general REChgT is and if REChgT remains effective in more challenging cases, we studied the relative stability of the Glu-Lys pair buried in the hydrophobic core of two SNase mutants, V23E/L36K (PDB ID: 3NHH[60]) and V23K/L36E (PDB ID: 6AMF[59]) of Δ+PHS. Non-replica exchange MSλD simulations were also performed as a comparison. As shown in Figure 5, there is insufficient sampling in alchemical space without replica exchange for both SNase mutants. At least two factors may contribute to this sampling issue: slow conformational relaxation and an imperfect $V_{bias}(\{\lambda\})$, the latter of which is optimized using ALF and we do not know how optimal it is in such challenging cases prior to any rigorous free energy calculations. These two problems have different origins and may need different strategies to be alleviated, but it is hard to separate their contributions solely based on the MSλD simulations. Compared to the above test case of DHOD/orotate where we performed another rigorous free energy calculation to obtain optimal $V_{bias}(\{\lambda\})$, the complications shown here is more frequently encountered and more general.

We first tried to directly apply REChgT by reducing partial charges of the Glu-Lys pair, assuming that slow structural transitions are the main cause of insufficient sampling. As shown in Figures S8 and S9, although improved sampling in alchemical space can be observed for both SNase mutants, it is still insufficient for well-converged free energy calculations within 20 ns. For instance, in the REChgT simulation of V23K/L36E (6AMF), Lys 23 sampled both protonated and unprotonated states, while Glu 36 was consistently trapped near the protonated state (Figure S8), implying that $V_{bias}(\{\lambda\})$ may not be sufficiently optimal to drive fast reversible transitions in alchemical space.

Instead of further optimizing $V_{bias}(\{\lambda\})$ using ALF, which has its own challenges if there is no easy way to obtain sufficient information in short trial simulations, we tried to further incorporate biasing potential replica exchange into REChgT, i.e., constructing a 2D replica exchange scheme. In one dimension, partial charges of residues related to the buried charge are reduced, and in the second dimension, $V_{bias}(\{\lambda\})$ is varied. Exchange is attempted between all nearest neighboring conditions (see Figure S10). Here, we assume that although $V_{bias}(\{\lambda\})$ from ALF may not be perfect, searching the biasing potential parameter space near the current $V_{bias}(\{\lambda\})$ using biasing potential replica exchange could find one condition that is able to sufficiently flatten the alchemical free energy landscape.[70] Also, charging free energy profiles can often be well approximated by the summation of two terms, $V_{fixed}(\{\lambda\})$ and $V_{quadratic}(\{\lambda\})$ (see Figure S1 for example), and further examination of the performance of ALF in charging free energy calculations for the orotate/DHOD system has indicated that ALF can reliably optimize $V_{quadratic}(\{\lambda\})$, but not $V_{fixed}(\{\lambda\})$ (see Figure S11). Therefore, $V_{fixed}(\{\lambda\})$ together with partial charges of the Glu-Lys pair are varied among different conditions in the 2D REChgT scheme. This design aims to alleviate the two aforementioned problems, slow conformational relaxation and imperfect $V_{bias}(\{\lambda\})$, at the same time. As shown in Figure 6, reversible transitions in alchemical space indeed could be sampled in these 2D REChgT simulations, a significant improvement over the original MSλD simulations. We then used WHAM to estimate the folding free energy difference when the Glu-Lys pair is changed from charge-neutral to ionized form. As shown in Figure 7, the results are

highly consistent with those obtained from windowing approach together with the MBAR estimator. Moreover, it reaches self consistency within ~ 60 ns. Such a high convergence rate was not observed in the MSλD simulations.

Protonation states of the buried Glu-Lys pair in SNase

The well-converged free energy results in the present study also allow us to further examine the protonation state of this Glu-Lys pair in the hydrophobic interior of SNase protein. A previous experimental study[59] has found that the free energy contribution due to the coupling between Lys23 and Glu36 in 6AMF is −2.4 kcal/mol at pH 7, which is similar to side chain HB strengths in proteins. Furthermore, it matches the coupling free energy for an analogous neutral pair (Lys-Gln) buried inside the same background protein. Therefore, it has been suggested that Lys23 and Glu36 prefer to be neutral when introduced to the SNase Δ+PHS protein. Based on Figure 7A, our simulation result is consistent with the experimental observation, and we predict that the neutral form is ~1 kcal/mol more stable than the charged form. Such preference of the Lys23-Glu36 pair being neutral inside SNase Δ+PHS has also been observed by previous TI simulations using the same force field.[11]

For the other Glu-Lys pair, Glu23-Lys36 in 3NHH, NMR spectroscopy has shown that the resonances of Glu23 are within the range normally observed for surface Glu residues.[60] Also, the coupling free energy between Glu23 and Lys36 was −5kcal/mol at pH 7,[60] much stronger than that for the Lys23-Glu36 pair in 6AMF (−2.4 kcal/mol). These observations suggest that Glu23-Lys36 are charged inside the the protein. Unfortunately, our simulations (and previous TI simulations using the same force field[11]) did not recapitulate the suggested interpretation. Multiple factors may contribute to the discrepancies between our simulation and the experimental results. Considering that we only used an additive force field and did not consider electronic polarization, which may be essential for stabilizing buried ion pairs in SNase mutant,[12] the level of agreement with experimental observations we observed here appears to be quite encouraging. Additionally, the interaction between a Ca^{2+} ion and Asp21 seem to be very important for stabilizing the $\beta 1 - \beta 2$ loop,[59] but we did not carefully optimize the parameters related to Ca^{2+} to properly recapitulate the dynamic structure of $\beta 1 - \beta 2$ loop. This may contribute to a systematic shift of $\Delta\Delta G_{fold}(N \rightarrow C)$ for both mutants.

Conclusion

Buried titratable groups often play important functional roles in many biological processes. Therefore, there is a great need to understand their properties and stabilizing mechanisms for both theoretical and practical purposes. This requires obtaining high resolution spatial and temporal information of the system, such as protonation states of the ionizable groups, protein conformational dynamics, and fluctuations of penetrating water molecules (if present), which pose great challenges for both experimental and computational studies. In the present study, we developed an enhanced sampling method named REChgT for MSλD free energy simulations of systems with buried charges. Two test systems were studied, DHOD in complex with orotate and two SNase mutants with a buried Glu-Lys pair. We found that slow convergence in alchemical space was a direct result of slow conformational rearrangement related to the buried charges. Considering that many local

strong interactions related to the buried charge are likely polar, such as HBs and salt bridges, the key idea of REChgT is to reduce relevant electrostatic interactions by reducing the partial charges, thus lowering the barrier of structural reorganization. In comparison with conventional MSλD simulations, the use of REChgT could significantly increase sampling in both conformational and alchemical space. This directly translates into faster convergence of free energy calculations. For instance, both test cases require tens of ns/replica to achieve self-consistency when REChgT is utilized, which is unprecedented compared to the unaccelerated MSλD simulations.

The dramatic improvement of sampling efficiency in REChgT simulations is expected to greatly benefit constant pH molecular dynamics simulations of complex systems, where pH could regulate the protonation states of titratable residues, thus modulating protein function. Note that MSλD is a high throughput free energy method and can be used to study multiple titratable groups simultaneously in a single simulation. Although more replicas are needed in REChgT simulations, the number of replicas required only depends on how much electrostatic interactions need to be reduced in order to allow for more structural reorganizations and how well ALF performs in optimizing biasing potentials, but not the number of titratable groups in a system. In the future, we may consider adding another phase of optimizing biasing potentials in ALF using short, trial 2D REChgT simulations (rather than the standard MSλD simulations in the current two phases of ALF). This is expected to give better biasing potentials for such challenging systems, which in return could require less replicas in the production simulations. Also, considering that all replicas are run in parallel in REChgT simulations, it is likely still advantageous to use more computational resources (not necessarily more wall time) to obtain improved sampling and faster convergence of free energy calculations. Similarly, relative free energy calculations for small molecules with net charge changes may be improved by REChgT as well, owing to its ability to improve local structural rearrangement. Furthermore, we note that in principle REChgT is not limited to MSλD simulations, and it can be applied to enhance sampling of other free energy methods, such as FEP and TI. However, since these methods scale poorly with the number of end states studied, additional design might be needed to reduce the computational cost. We will explore the applicability of REChgT in our future work.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

We gratefully acknowledge funding from the NIH (GM130587).

References

(1). Fersht A, et al. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding; Macmillan, 1999.

(2). Harris TK; Turner GJ Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. IUBMB life 2002, 53, 85–98. [PubMed: 12049200]

(3). Warshel A Energetics of enzyme catalysis. Proceedings of the National Academy of Sciences 1978, 75, 5250–5254.

(4). Cannon WR; Benkovic SJ Solvation, reorganization energy, and biological catalysis. Journal of Biological Chemistry 1998, 273, 26257–26260. [PubMed: 9756847]

(5). Dyla M; Kjærgaard M; Poulsen H; Nissen P Structure and mechanism of P-type ATPase ion pumps. Annual review of biochemistry 2020, 89, 583–603.

(6). Wikstrom M; Sharma V; Kaila VR; Hosler JP; Hummer G New perspectives on proton pumping in cellular respiration. Chemical reviews 2015, 115, 2196–2221. [PubMed: 25694135]

(7). Kaila VR Resolving chemical dynamics in biological energy conversion: Long-range proton-coupled electron transfer in respiratory complex I. Accounts of Chemical Research 2021, 54, 4462–4473. [PubMed: 34894649]

(8). Nicholls DG; Ferguson SJ Bioenergetics, 4th ed.; Academic Press, Elsevier: Amsterdam, 2013; pp xiv, 419 pages.

(9). Von Ballmoos C; Wiedenmann A; Dimroth P Essentials for ATP synthesis by F1F0 ATP synthases. Annual review of biochemistry 2009, 78, 649–672.

(10). Iwata S; Lee JW; Okada K; Lee JK; Iwata M; Rasmussen B; Link TA; Ramaswamy S; Jap BK Complete structure of the 11-subunit bovine mitochondrial cytochrome bc1 complex. Science 1998, 281, 64–71. [PubMed: 9651245]

(11). Deng J; Cui Q Reverse Protonation of Buried Ion-Pairs in Staphylococcal Nuclease Mutants. Journal of Chemical Theory and Computation 2021, 17, 4550–4563. [PubMed: 34143626]

(12). Deng J; Cui Q Electronic Polarization Is Essential for the Stabilization and Dynamics of Buried Ion Pairs in Staphylococcal Nuclease Mutants. Journal of the American Chemical Society 2022, 144, 4594–4610. [PubMed: 35239338]

(13). Torabifard H; Panahi A; Brooks CL III M2 amphipathic helices facilitate pH-dependent conformational transition in influenza A virus. Proceedings of the National Academy of Sciences 2020, 117, 3583–3591.

(14). Clark AJ; Negron C; Hauser K; Sun M; Wang L; Abel R; Friesner RA Relative binding affinity prediction of charge-changing sequence mutations with FEP in protein–protein interfaces. Journal of molecular biology 2019, 431, 1481–1493. [PubMed: 30776430]

(15). Goh GB; Laricheva EN; Brooks CL III Uncovering pH-dependent transient states of proteins with buried ionizable residues. Journal of the American Chemical Society 2014, 136, 8496–8499. [PubMed: 24842060]

(16). Khandogin J; Brooks CL III Toward the accurate first-principles prediction of ionization equilibria in proteins. Biochemistry 2006, 45, 9363–9373. [PubMed: 16878971]

(17). Liu P; Kim B; Friesner RA; Berne BJ Replica exchange with solute tempering: A method for sampling biological systems in explicit water. Proceedings of the National Academy of Sciences of the United States of America 2005, 102, 13749–13754. [PubMed: 16172406]

(18). Wang L; Friesner RA; Berne B Replica exchange with solute scaling: a more efficient version of replica exchange with solute tempering (REST2). The Journal of Physical Chemistry B 2011, 115, 9431–9438. [PubMed: 21714551]

(19). Wang L; Berne B; Friesner RA On achieving high accuracy and reliability in the calculation of relative protein–ligand binding affinities. Proceedings of the National Academy of Sciences 2012, 109, 1937–1942.

(20). Kamiya M; Sugita Y Flexible selection of the solute region in replica exchange with solute tempering: Application to protein-folding simulations. The Journal of chemical physics 2018, 149.

(21). Jung J; Kobayashi C; Sugita Y Acceleration of generalized replica exchange with solute tempering simulations of large biological systems on massively parallel super-computer. Journal of Computational Chemistry 2023, 44, 1740–1749. [PubMed: 37141320]

(22). Ross GA; Bodnarchuk MS; Essex JW Water sites, networks, and free energies with grand canonical Monte Carlo. Journal of the American Chemical Society 2015, 137, 14930–14943. [PubMed: 26509924]

(23). Ross GA; Bruce Macdonald HE; Cave-Ayland C; Cabedo Martinez AI; Essex JW Replica-exchange and standard state binding free energies with Grand Canonical Monte Carlo. Journal of Chemical Theory and Computation 2017, 13, 6373–6381. [PubMed: 29091438]

(24). Ross GA; Russell E; Deng Y; Lu C; Harder ED; Abel R; Wang L Enhancing water sampling in free energy calculations with grand canonical Monte Carlo. Journal of Chemical Theory and Computation 2020, 16, 6061–6076. [PubMed: 32955877]

(25). Ge Y; Melling OJ; Dong W; Essex JW; Mobley DL Enhancing sampling of water rehydration upon ligand binding using variants of grand canonical Monte Carlo. Journal of Computer-Aided Molecular Design 2022, 36, 767–779. [PubMed: 36198874]

(26). Miao Y; McCammon JA Annual reports in computational chemistry; Elsevier, 2017; Vol. 13; pp 231–278. [PubMed: 29720925]

(27). Miao Y; Feher VA; McCammon JA Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. Journal of chemical theory and computation 2015, 11, 3584–3595. [PubMed: 26300708]

(28). Wang J; Arantes PR; Bhattarai A; Hsu RV; Pawnikar S; Huang Y.-m. M.; Palermo G; Miao Y Gaussian accelerated molecular dynamics: Principles and applications. Wiley Interdisciplinary Reviews: Computational Molecular Science 2021, 11, e1521. [PubMed: 34899998]

(29). Miao Y; Bhattarai A; Wang J Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. Journal of chemical theory and computation 2020, 16, 5526–5547. [PubMed: 32692556]

(30). Zuckerman DM; Chong LT Weighted ensemble simulation: review of methodology, applications, and software. Annual review of biophysics 2017, 46, 43–57.

(31). Ray D; Gokey T; Mobley DL; Andricioaei I Kinetics and free energy of ligand dissociation using weighted ensemble milestoning. The Journal of chemical physics 2020, 153,154117. [PubMed: 33092382]

(32). Roussey NM; Dickson A Enhanced Jarzynski free energy calculations using weighted ensemble. The Journal of Chemical Physics 2020, 153, 134116. [PubMed: 33032408]

(33). Laio A; Parrinello M Escaping free-energy minima. Proceedings of the national academy of sciences 2002, 99, 12562–12566.

(34). Bussi G; Laio A Using metadynamics to explore complex free-energy landscapes. Nature Reviews Physics 2020, 2, 200–212.

(35). Bonomi M; Barducci A; Parrinello M Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics. Journal of computational chemistry 2009, 30, 1615–1621. [PubMed: 19421997]

(36). Bertazzo M; Gobbo D; Decherchi S; Cavalli A Machine learning and enhanced sampling simulations for computing the potential of mean force and standard binding free energy. Journal of chemical theory and computation 2021, 17, 5287–5300. [PubMed: 34260233]

(37). Zheng L; Chen M; Yang W Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. Proceedings of the National Academy of Sciences 2008, 105, 20227–20232.

(38). Zheng L; Chen M; Yang W Simultaneous escaping of explicit and hidden free energy barriers: Application of the orthogonal space random walk strategy in generalized ensemble based conformational sampling. The Journal of chemical physics 2009, 130.

(39). Zheng L; Yang W Practically efficient and robust free energy calculations: Double-integration orthogonal space tempering. Journal of chemical theory and computation 2012, 8, 810–823. [PubMed: 26593343]

(40). Lu C; Li X; Wu D; Zheng L; Yang W Predictive sampling of rare conformational events in aqueous solution: Designing a generalized orthogonal space tempering method. Journal of chemical theory and computation 2016, 12, 41–52. [PubMed: 26636477]

(41). Lyubartsev A; Martsinovski A; Shevkunov S; Vorontsov-Velyaminov P New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. The Journal of chemical physics 1992, 96, 1776–1783.

(42). Zhang S; Hahn DF; Shirts MR; Voelz VA Expanded ensemble methods can be used to accurately predict protein-ligand relative binding free energies. Journal of chemical theory and computation 2021, 17, 6536–6547. [PubMed: 34516130]

(43). Monroe JI; Shirts MR Converging free energies of binding in cucurbit [7] uril and octa-acid host–guest systems from SAMPL4 using expanded ensemble simulations. Journal of computer-aided molecular design 2014, 28, 401–415. [PubMed: 24610238]

(44). Miao Y; McCammon JA Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. Molecular simulation 2016, 42, 1046–1055. [PubMed: 27453631]

(45). Liao Q Enhanced sampling and free energy calculations for protein simulations. Progress in molecular biology and translational science 2020, 170, 177–213. [PubMed: 32145945]

(46). Haochuan C; Haohao F; Xueguang S; Wensheng C Importance sampling methods and free energy calculations. Progress in Chemistry 2018, 30, 921.

(47). Valsson O; Tiwary P; Parrinello M Enhancing important fluctuations: Rare events and metadynamics from a conceptual viewpoint. Annual review of physical chemistry 2016, 67, 159–184.

(48). Comitani F; Gervasio FL Modeling Ligand–Target Binding with Enhanced Sampling Simulations. Biomol. Simul. Struct. Based Drug Discov 2019, 45–66.

(49). Fu H; Shao X; Cai W; Chipot C Taming rugged free energy landscapes using an average force. Accounts of chemical research 2019, 52, 3254–3264. [PubMed: 31680510]

(50). Kong X; Brooks CL III λ-dynamics: A new approach to free energy calculations. The Journal of chemical physics 1996, 105, 2414–2423.

(51). Knight JL; Brooks CL III Multisite λ dynamics for simulated structure–activity relationship studies. Journal of chemical theory and computation 2011, 7, 2728–2739. [PubMed: 22125476]

(52). Hayes RL; Armacost KA; Vilseck JZ; Brooks CL III Adaptive landscape flattening accelerates sampling of alchemical space in multisite λ dynamics. The Journal of Physical Chemistry B 2017, 121, 3626–3635. [PubMed: 28112940]

(53). Hayes RL; Vilseck JZ; Brooks CL III Approaching protein design with multisite λ dynamics: Accurate and scalable mutational folding free energies in T4 lysozyme. Protein Science 2018, 27, 1910–1922. [PubMed: 30175503]

(54). Kumar S; Rosenberg JM; Bouzida D; Swendsen RH; Kollman PA The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. Journal of computational chemistry 1992, 13, 1011–1021.

(55). Shirts MR; Chodera JD Statistically optimal analysis of samples from multiple equilibrium states. The Journal of chemical physics 2008, 129, 124105. [PubMed: 19045004]

(56). Nørager S; Jensen KF; Björnberg O; Larsen SE coli dihydroorotate dehydrogenase reveals structural and functional distinctions between different classes of dihydroorotate dehydrogenases. Structure 2002, 10, 1211–1223. [PubMed: 12220493]

(57). Feig M; Karanicolas J; Brooks CL III MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. Journal of Molecular Graphics and Modelling 2004, 22, 377–395. [PubMed: 15099834]

(58). Vilseck JZ; Cervantes LF; Hayes RL; Brooks CL III Optimizing Multisite λ-Dynamics Throughput with Charge Renormalization. Journal of chemical information and modeling 2022, 62, 1479–1488. [PubMed: 35286093]

(59). Robinson AC; Schlessman JL; García-Moreno EB Dielectric properties of a protein probed by reversal of a buried ion pair. The Journal of Physical Chemistry B 2018, 122, 2516–2524. [PubMed: 29466010]

(60). Robinson AC; Castañeda CA; Schlessman JL; García-Moreno EB Structural and thermodynamic consequences of burial of an artificial ion pair in the hydrophobic interior of a protein. Proceedings of the National Academy of Sciences 2014, 111, 11685–11690.

(61). Olsson MH; Søndergaard CR; Rostkowski M; Jensen JH PROPKA3: consistent treatment of internal and surface residues in empirical p K a predictions. Journal of chemical theory and computation 2011, 7, 525–537. [PubMed: 26596171]

(62). Søndergaard CR; Olsson MH; Rostkowski M; Jensen JH Improved treatment of ligands and coupling effects in empirical calculation and rationalization of p K a values. Journal of chemical theory and computation 2011, 7, 2284–2295. [PubMed: 26606496]

(63). Buckner J; Liu X; Chakravorty A; Wu Y; Cervantes L; Lai T; Brooks CL III pyCHARMM: Embedding CHARMM Functionality in a Python Framework. 2023,

(64). Hayes RL; Buckner J; Brooks CL III BLaDE: A basic lambda dynamics engine for GPU-accelerated molecular dynamics free energy calculations. Journal of chemical theory and computation 2021, 17, 6799–6807. [PubMed: 34709046]

(65). Huang J; Rauscher S; Nawrocki G; Ran T; Feig M; de Groot BL; Grubmuller H; MacKerell AD CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nature Methods 2017, 14, 71–73. [PubMed: 27819658]

(66). Darden T; York D; Pedersen L Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. The Journal of chemical physics 1993, 98, 10089–10092.

(67). Van Gunsteren W; Berendsen HJ Algorithms for macromolecular dynamics and constraint dynamics. Molecular Physics 1977, 34, 1311–1327.

(68). Ding X; Vilseck JZ; Brooks CL III Fast solver for large scale multistate Bennett acceptance ratio equations. Journal of chemical theory and computation 2019, 15, 799–802. [PubMed: 30689377]

(69). Ding X; Vilseck JZ; Hayes RL; Brooks CL III Gibbs sampler-based λ-dynamics and Rao–Blackwell estimator for alchemical free energy calculation. Journal of chemical theory and computation 2017, 13, 2501–2510. [PubMed: 28510433]

(70). Armacost KA; Goh GB; Brooks CL III Biasing potential replica exchange multisite λ-dynamics for efficient free energy calculations. Journal of chemical theory and computation 2015, 11, 1267–1277. [PubMed: 26579773]

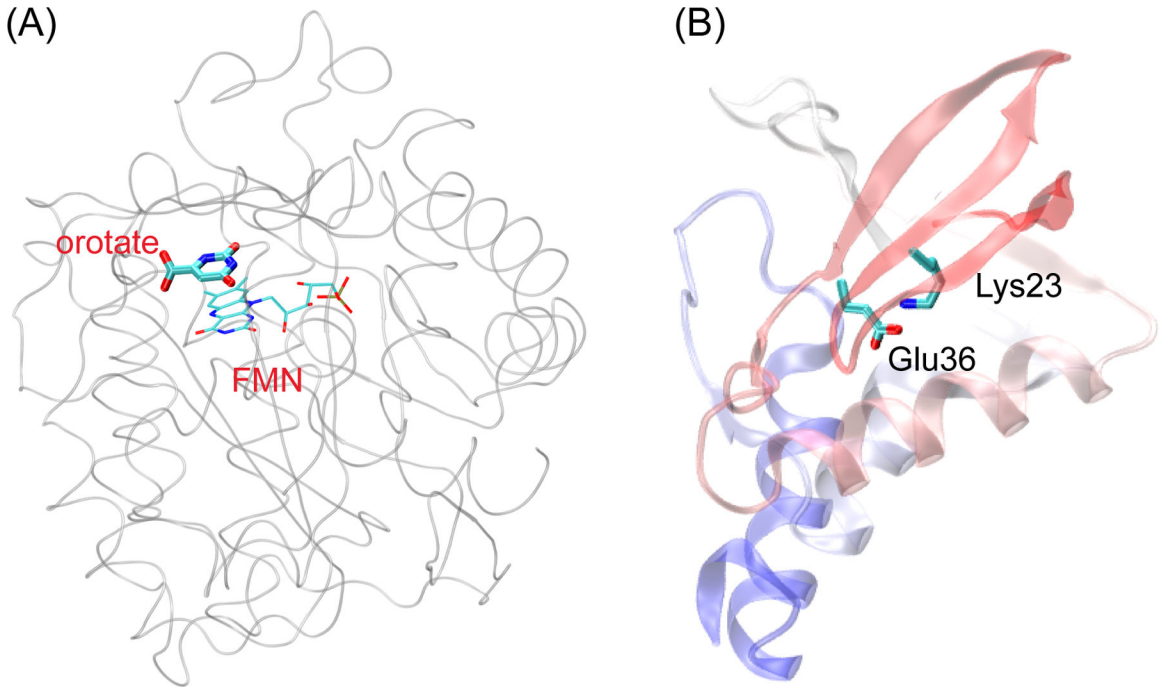


Figure 1:
(A) Experimental structure of orotate in complex with protein DHOD (PDB: 1F76). The protein is shown in grey. Molecule orotate and cofactor FMN are labeled. (B) Experimental structure of SNase Δ+PHS variant with Lys23 and Glu36 buried inside the hydrophobic core (PDB: 6AMF). The protein is shown in cartoon, with color changing from red at the N-terminus to blue at the C-terminus. The other SNase variant studied in this work has Glu23 and Lys36 buried inside the hydrophobic core (PDB: 3NHH). The structure is not shown here.

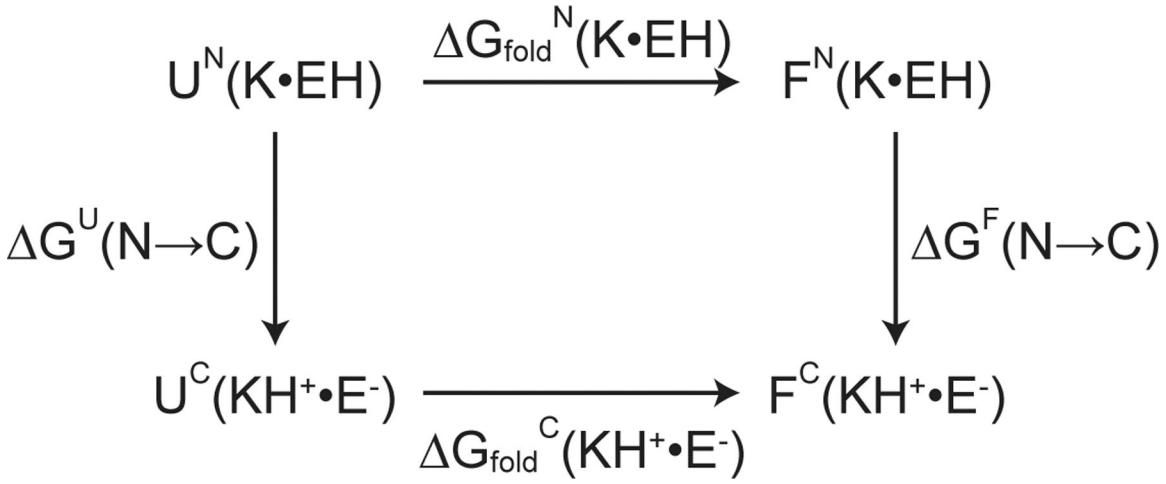


Figure 2:
Thermodynamic cycle used to estimate the change in folding free energy when the Glu-Lys pair changes from the neutral (N) form to the charged form (C). U represents the unfolded state and F represents the folded state. Our free energy simulations consider the two vertical processes.

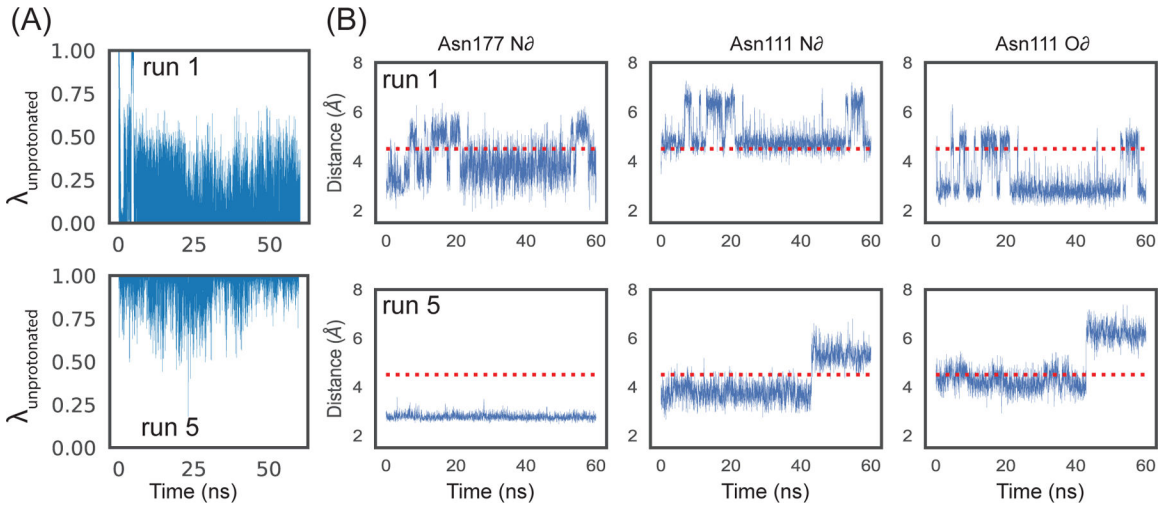


Figure 3:

Limited sampling of two representative MSλD simulations of orotate in complex with protein DHOD. The first row corresponds to one simulation (run 1) and the second row corresponds to another simulation (run 5). See Figures S2 and S3 for plots for all other independent runs of MSλD simulations. (A) λ of the unprotonated state of orotate as a function of simulation time. (B) Distances between one of the oxygen atoms in the carboxylate group of orotate and three atoms in the neighboring residues. The red line indicates 4.5 Å in all cases.

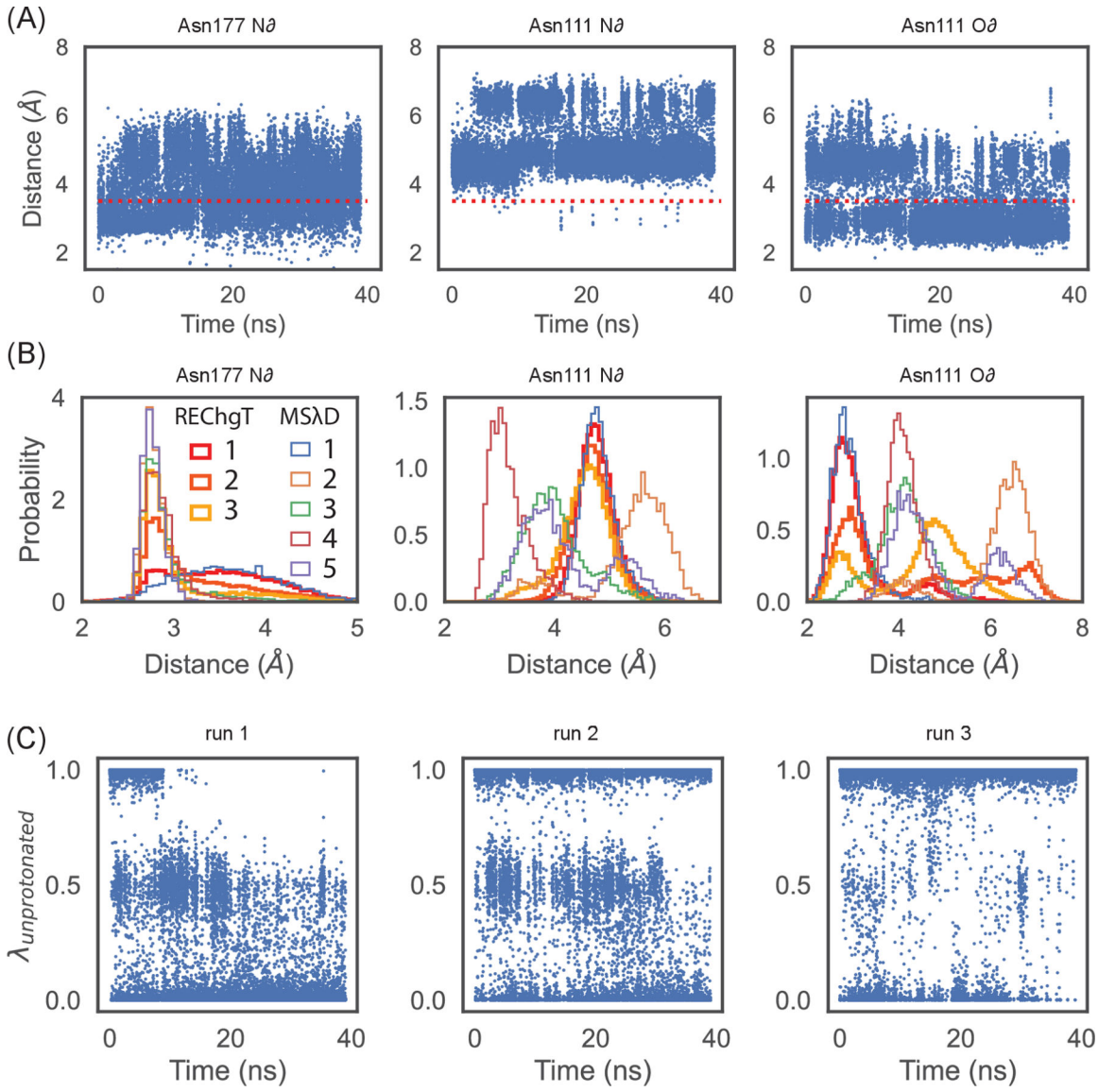


Figure 4:

Improved sampling in both conformational and alchemical space of REChgT simulations of orotate in complex with protein DHOD. (A) Distances between one of the oxygen atoms in the carboxylate group of orotate and three atoms in the neighboring residues in one representative REChgT simulation. The red line indicates 4.5 Å in all cases. See Figure S5 for plots from all independent REChgT simulations. (B) Probability distribution of minimum distances between the oxygen atoms in the orotate carboxylate group and three atoms in the neighboring residues in all MSλD and REChgT simulations. (C) λ of the unprotonated state of orotate under the condition of unscaled partial charges as a function of simulation time in three REChgT simulations.

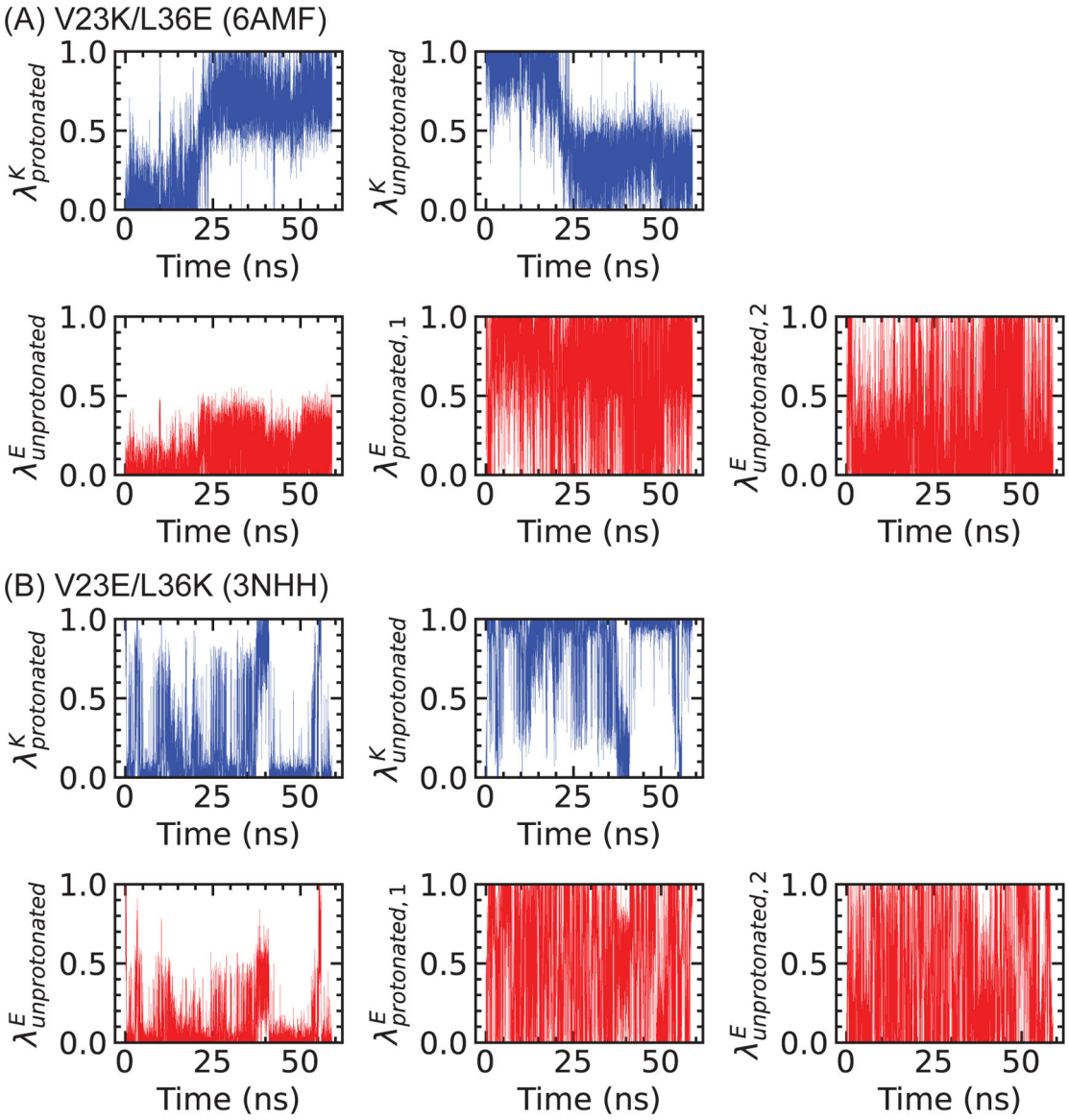


Figure 5:

Representative plots of λ's for the Glu-Lys pair as a function of simulation time in original MSλD simulations of two SNase mutants: (A) V23K/L36E (PDB ID: 6AMF) and (B) V23E/L36K (PDB ID: 3NHH). See Figures S6 and S7 for plots from all independent MSλD simulations.

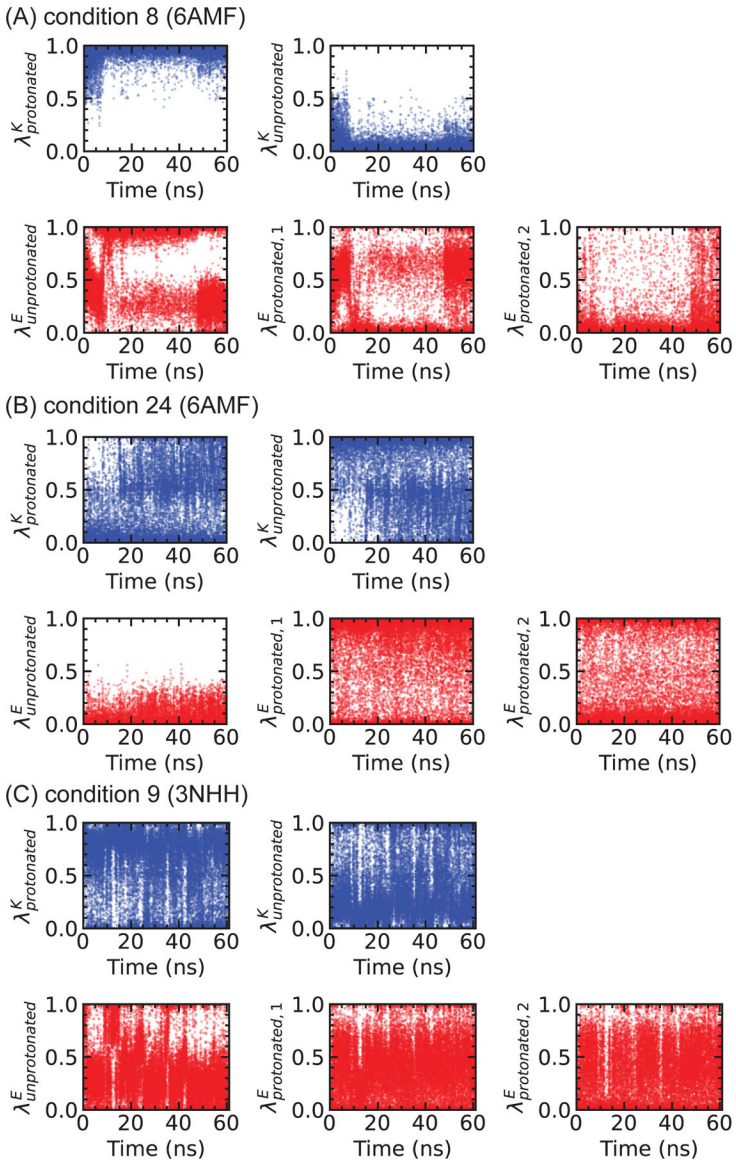


Figure 6:

λ's for the Glu-Lys pair as a function of simulation time in the 2D REChgT simulation of V23K/L36E (6AMF) under conditions 8 (A) and 24 (B) and 2D REChgT simulation of V23E/L36K (3NHH) under condition 9 (C). For better visual comparison with Figure 5, only the first 60 ns data were shown.

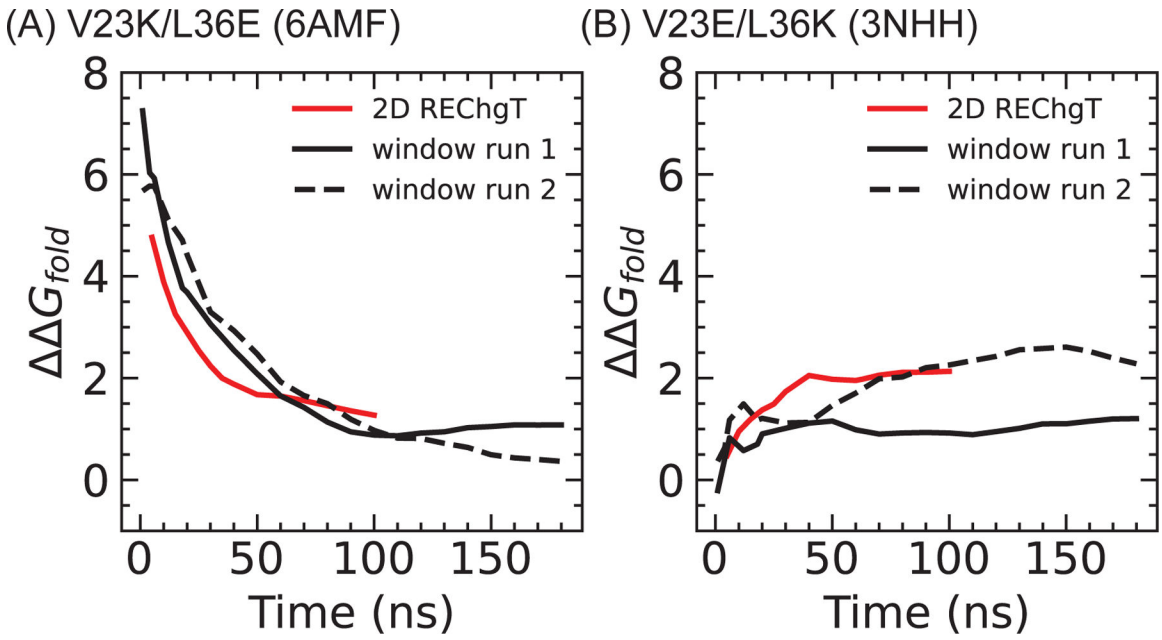


Figure 7:

Change in protein stability when the Glu-Lys pair changes from neutral form to ionized form ($\Delta\Delta G_{fold}$) as a function of simulation time (per replica) for (A) V23K/L36E (6AMF) and (B) V23E/L36K (3NHH) based on the 2D REChgT simulations (red traces) and the windowing approach (black traces).

Table 1:

Comparison of free energy results obtained from REChgT simulations and the FEP-like windowing approach together with the MBAR estimator. Here the free energy difference for orotate changing from unprotonated state to protonated state in DHOD-bound state ($\Delta G_{protonation}$) is reported.

Protocol	run	$\Delta G_{protonation}$(kcal/mol)	$\langle \Delta G_{protonation} \rangle$(kcal/mol)
REChgT	1	29.06	30.15 ± 0.78
	2	30.53	
	3	30.86	
windowing/MBAR	1	30.70	30.20 ± 0.36
	2	30.03	
	3	29.88	